

WHAT IS CLAIMED IS:

1. A method of source normalization training for HMM modeling of speech comprising the steps of:

5 (a) providing an initial model;
(b) on said initial model or following new models performing the following steps to get a new model:

b₁) estimation of intermediate quantities;

10 b₂) performing re-estimation to determine initial state probability, transition probability, mixture component probability and environment probability;

b₃) deriving mean vector and bias vector;

b₄) solving jointly for mean vector and bias vector using linear equations and determining variances and transformation; and

15 b₅) replacing old model parameters for the calculated ones; and

c) determining after a new model is formed if it differs significantly from the previous model and if so repeating steps b₁ - b₅.

2. The method of Claim 1 wherein in step b₁ estimation intermediate quantities is determined by $\alpha_i(j, e) \triangleq p(o'_i, \theta_i = j, \varphi = e | \bar{\lambda})$,

$\beta_i(j, e) \triangleq p(o_{i+1}^T | \theta_i = j, \varphi = e, \bar{\lambda})$, and $\gamma_i(j, k, e) \triangleq p(\theta_i = j, \xi = k, \varphi = e | O, \bar{\lambda})$.

3. The method of Claim 2 wherein step b₂ the initial state probability is

determined by $u_i = \frac{1}{R} \sum_{r=1}^R \frac{\sum_{e \in \Omega_e} \alpha'_i(i, e) \beta'_i(i, e)}{\sum_{i \in \Omega_i} \sum_{e \in \Omega_e} \alpha'_i(i, e) \beta'_i(i, e)}$, transition probability is determined by

$$a_{ij} = \frac{\bar{a}_{ij} \sum_{r=1}^R \frac{1}{p(O^r | \bar{\lambda})} \sum_{e \in \Omega_e} \sum_{t=1}^{T'} \alpha'_i(i, e) b_{j,e}(O'_{t+1}) \beta'_{t+1}(j, e)}{\sum_{r=1}^R \frac{1}{p(O^r | \bar{\lambda})} \sum_{e \in \Omega_e} \sum_{t=1}^{T'} \alpha'_i(i, e) \beta'_i(i, e)}, \text{ mixture component probability is}$$

determined by $c_{jk} = \frac{\sum_{r=1}^R \sum_{e \in \Omega_e} \sum_{t=1}^{T'} \gamma'_t(j, k, e)}{\sum_{r=1}^R \frac{1}{p(O^r | \bar{\lambda})} \sum_{e \in \Omega_e} \sum_{t=1}^{T'} \alpha'_i(j, e) \beta'_i(j, e)}$, and environment

probability is determined by $l_e = \frac{1}{R} \sum_{r=1}^R \frac{\sum_{i \in \Omega_i} \alpha'_i(j, e)}{\sum_{e \in \Omega_e} \sum_{i \in \Omega_i} \alpha'_i(j, e)}$.

4. The method of Claim 2 wherein step b₃ deriving mean vector and bias

vector is determined by $\rho(j, k, e) \triangleq \sum_{r=1}^R \sum_{t=1}^{T'} \gamma'_t(j, k, e) O'_t$; $g(j, k, e) \triangleq \sum_{r=1}^R \sum_{t=1}^{T'} \gamma'_t(j, k, e)$,

$$G_{ke} = \sum_{j \in \Omega_j} g(j, k, e) \sum_{jk}^{-1}, \quad E_{jke} = g(j, k, e) W_{je}' \sum_{jk}^{-1}, \quad F_{jk} = \sum_{e \in \Omega_e} E_{jke} W_{je},$$

$$a_{jk} = \sum_{e \in \Omega_e} W_{je}' \sum_{jk}^{-1} \rho(j, k, e), \text{ and } c_{ke} = \sum_{j \in \Omega_j} \sum_{jk}^{-1} \rho(j, k, e).$$

5. The method of Claim 2 wherein step b₄ equations

$$\sum_{e \in \Omega_e} E_{jke} b_{ke} + F_{jk} \mu_{jk} = a_{jk} \quad \forall j \in \Omega_j, \text{ and } G_{ke} b_{ke} + \sum_{j \in \Omega_j} H_{jke} \mu_{jk} = c_{ke} \quad \forall e \in \Omega_e \text{ are used}$$

for solving jointly and equation $\sum_{jk} = \frac{\sum_{e \in \Omega_e} \sum_{r=1}^R \sum_{t=1}^{T'} \gamma'_t(j, k, e) \delta'_t(j, k, e) \delta'_t(j, e, k)'}{\sum_{e \in \Omega_e} g(j, k, e)}$

is used to determine variance and equations $Z_{je}^{(m)} = W_{je}^{(m)} R_{je}(m)$,

$$Z_{je}^{(m,n)} \triangleq \sum_{k \in \Omega_m} \sum_{jk}^{-l(m,n)} \mu_{jk}^{(n)} \sum_{r=1}^R \sum_{t=1}^T \gamma_t^r(j, k, e) (o_t^r - b_{ke})^{(m)},$$

and

$$R_{je}^{(p,n)}(m) \triangleq \sum_{k \in \Omega_m} \sum_{jk}^{-l(m,n)} \mu_{jk}^{(p)} \mu_{jk}^{(n)} \sum_{r=1}^R \sum_{t=1}^T \gamma_t^r(j, k, e).$$

are used to determine transformation.

5

6. An improved speech recognition system comprising:

a speech recognizer; and

a source normalization model derived by application of an estimation maximization algorithm.

SUB
Q2

ADD
Q3

add
B1
add E1